

雑談を行うことでユーザの印象を向上させる インタビュー対話システム

Survey Dialogue Systems that Improves Users Impressions using Small Talks

小堀嵩博*¹ 中野幹生*² 中村友昭*¹
Takahiro Kobori Mikio Nakano Tomoaki Nakamura

*¹電気通信大学
The University of Electro-Communications

*²(株)ホンダ・リサーチ・インスティテュート・ジャパン
Honda Research Institutes Japan Co., Ltd.

Previously build survey dialogue systems mainly focused on obtaining information from the users, thus they just iterated asking questions. However, users might not want to repeatedly use such systems. This paper proposes a method for dialogue management for engaging in natural conversation using small talks while it performs survey dialogues based on frame-based dialogue management. Experimental results showed the proposed method gives better impression to the users than survey dialogues without small talks.

1. はじめに

我々は、ユーザから情報を聞き出してその情報を構造化する対話システム、すなわち、インタビューを行う対話システムの構築を目指している。そのようなシステムの一例として、ユーザの食生活をサポートするために、ユーザが何を食べたのか聞いて記録するようなシステムが考えられる。本稿では、そのようなシステムの対話管理手法を提案する。

従来のインタビュー対話システムの対話管理は、主にユーザの情報を得ることに主眼を置いていたため、質問を繰り返すだけの機械的な対話を行っていた。一度きりのサーベイならそれでも良いが、ユーザは対話を楽しむことができなければ、継続的に使おうとはしないと考えられるので、毎日の食事内容を記録するようなシステムには向いていないと考えられる。

人間同士の対話を観察すると、対話を楽しむ雑談の中で、相手から情報を聞き出すような場面が見受けられる。このような対話を実現できれば、楽しんで使ってもらえるインタビュー対話システムができると考えられる。しかしながら、現状の雑談対話システムの性能はまだ不十分であり、雑談を主体にする、対話が破たんしてしまう [Higashinaka 15]。

そこで、本稿では、インタビューを主体にしながら、時々潤滑発話を行う対話管理方法を提案する。潤滑発話とは、対話を円滑にする効果のある発話のことであり、相槌、自己開示などが例として考えられる。潤滑発話を生成することでユーザが対話を楽しむことができ、ユーザは対話システムをまた使いたいと考えてくれると思われる。

2. 先行研究

従来のインタビューを行う対話システムは主にユーザの情報を正確に得ることに主眼を置いていた [Gary 85, Stent 06, Johnston 13]。[Stent 06]ではプロトタイプアンケートを行う対話システムを提案しており、大学のコース評価に適用している。[Johnston 13]ではアメリカの一般的な社会調査の質問をいくつか聞く対話システムを構築している。しかし、これらのシステムは事務的な対話しかできないので、ユーザが使っていて楽しいと思えない可能性が高い。したがって、仕事ではない場面で気軽に使い続けてもらうのは難しいのではないかと考えられる。

これらのシステムとは逆に、非タスク型の対話システムの研究 [磯村 09, 東中 14]では、ユーザが対話を楽しめるようなシステムが研究されている。

本研究では、この2つの考えを組み合わせインタビューの合間に雑談を挟むことでユーザに良い印象を与えつつ、ユーザの情報を取得できる対話システムの構築法を提案する。

3. 提案手法

人同士のインタビュー対話では、質問のみを一時的に行いそれに答えていくのではなく、聞きたい内容とは直接的には関係のない雑談を話していることが多い。さらに、タスク指向型対話システムにおいてもユーザの半数は雑談を入力しているという結果がある [東中 14]。このことから、本研究では潤滑発話を質問の合間に用いることを提案する。潤滑発話を入れることでより人に近いインタビューを行うことができ、ユーザの印象が良くなる効果が見込まれる。

我々は、潤滑発話の有用性を調べるために事前にアンケートを行った。このアンケートでは質問の合間に潤滑発話をはさんだ発話数(質問のみ、潤滑発話1回、2回、3回、潤滑発話連続2回、3回)により人間の評価がどのように変化するのか調査した。ここで提示した対話例は人が考えて作ったものであり、誤解をすることなく完全な対話が出来ている。このアンケートの結果、質問の合間に3回潤滑発話を挟んだ対話例が最も支持を集めており、インタビューの合間に潤滑発話を多く発話した方がユーザに与える印象が良いと考えられる。しかし、雑談発話はユーザがどのような発話をするのか予測できないため、ユーザの発話意図を理解することが難しい。そのため、人のようにインタビューの多くを雑談で占めるのではなく、質問を主体として雑談を質問の合間に少し挟む方が良いと考えられる。

そこで、我々は、フレームベースの対話管理に基づいてインタビューを行いながら、潤滑発話を質問の合間に挟む対話管理手法を提案する。この方法を用いることで質問の合間に任意のタイミングで潤滑発話を挟むことが可能となる。

4. 実装:食事内容を聞く対話システム

提案手法に基づいて、食事内容を聞く対話システムを構築した。図1に本システムの処理の流れを示す。対話システムはユーザの発話を理解する言語理解部とシステムの対話を制御する対話管理部に分けることが出来る。

4.1 対話知識

対話システムはインタビューを行う上で必要となる知識をデータベースとして保持している。保持しているデータベースは表1のとおりである。潤滑発話、食事内容の質問については事前に行ったアンケートを用いて多くの発話例を収集し、対話システムが似たような発話を行わないようにした。

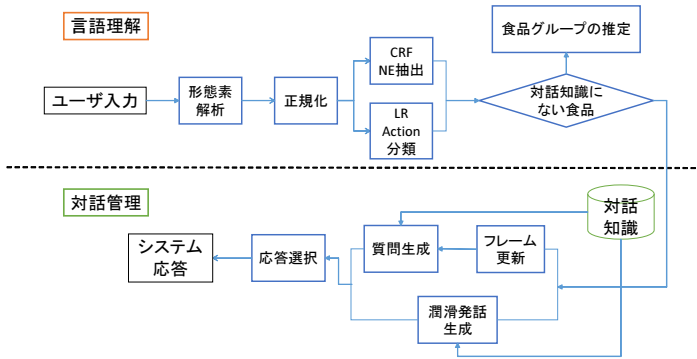


図 1: 実装した対話システムの処理の流れ

表 1: 対話知識の一覧

知識の種類	知識例
一般的な食事構成	朝: 主食, 主采, 汁物
食品グループに含まれる食品	主食: ご飯, パン, 麺, ...
食事内容の質問	朝は何を食べましたか?
潤滑発話	それ私の好物です!
食事の言い方	朝ご飯: 朝, 朝ごはん, ...
量に関する言い方	多い: たくさん, めっちゃ, ...

4.2 インタビューのための対話管理

システムは、フレームを用いた質問の対話管理を行っている。フレームに用いる属性として食事構成と食事の量を用いた。なお、食事構成については時間ごとに一般的と考えられる食事内容を食品グループ（主食、主菜、汁物等）を用いて定義している。そのため、ユーザが飲食した食品がどのグループなのか分からないと、フレームのどこに入れるかわからないため、食品グループを決定する必要がある。この判断はあらかじめ手作業で構築した知識を用いて行っているが、知識内に含まれていない食品もユーザ発話には存在するため、食品グループの推定を行っている。食品グループの推定については4.4節で述べる。図2が質問内容のフレーム例である。この例ではユーザが朝に“ご飯”、“納豆”、“味噌汁”を食べたと答えたときのフレームの内容になっている。

システムは朝食から時系列順に食事内容を聞いていく。これは、事前に収集した人同士のインタビュー対話においてほとんどのインタビューが時系列順に食事内容を聞いていたためである。また、質問は最初に大まかな食事内容を聞き、徐々に細かい内容を聞いていくように生成される。例えば、朝食に関する質問では「朝は何を食べました?」「他に何か食べました?」「主食は何でしたか?」というように質問を行っていく。この食事内容を聞いていく過程で事前に設定した食事構成のフレームが全て埋まった場合、その時点で食事構成に関する質問を終了し、食事の量を質問する。また、ユーザが何も食べていないと答えた場合は次の時間の食事について質問を行う。これを設定したすべての食事の時間で聞いていき、すべてのフレームが埋まり次第インタビューは終了となる。また、質問は毎回決まった言い回しではなく、対話知識の質問の中からランダムに選択される。

4.3 ユーザ発話の理解

本システムでは、ユーザ発話を理解するためにCRF (Conditional Random Fields) によるユーザ発話からの意味的な情報の抽出とLR (Logistic Regression) によるユーザ発話タイプの識別を用いて解析を行う。

CRFは系列ラベリング問題を解く手法であり、系列に対してそれに対応するラベルを推定する。本稿では、ユーザ発話の形態素の並びを入力系列とし、抽出したい情報をIOB2タグで

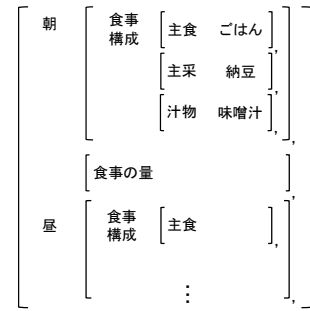


図 2: 質問内容のフレーム例

表し、それを出力ラベルとする。この形態素に対して特長を表す素性関数を設定しその素性関数の重要度を表す重みをCRFは学習する。ここでの抽出する情報は5種類（食品、食材、食品グループ、食事の量、食事の時間）である。本稿では、CRFの実装としてCRFsuite^{*1}を用いた。

発話タイプ推定のためのLRに用いる素性は語順を考慮することなく単語動詞の共起関係を示すBag of Words (BoW)を用いた。ユーザ発話タイプは3つ（挨拶、肯定、否定）である。本稿では、LRの実装としてLIBLINEAR[Fan 08]を用いた。

4.4 食品グループの推定

本稿の対話システムは食品グループを用いてインタビューの対話管理を行っているため、ユーザ情報から得られた食品がどの食品グループに当てはまるのか推定する必要がある。精度良く食品グループを推定するために、大塚ら [Otsuka 13] の手法を用いた。

食品グループの推定では、入力された食品名のBoW、文字の1-gram, 2-gram, 文字タイプ（ひらがな, カタカナ, 漢字, 英語）を素性とし、LRにより推定を行った。推定する食品グループは計8種類である（お菓子, 汁物, デザート, 主采, 一品物, 飲み物, 主食, 添え物）。

また、単純に食品グループの推定結果の1位のものをシステム応答として聞き返すのではなく、 n -bestを考慮し n 択で食品グループを聞き返すようにした。例えば、2択であれば「リングはお菓子とデザートのどちらですか?」というようにシステムは質問を行う。これは、必ずしもLRによる推定結果の1位のものが正解とは限らないためである。また、いくつか候補を出している場合ユーザはその中から食品グループの選択をするため、システムがより認識しやすい表現を入力してもらえると利点がある。ここでの、何択の質問をするかは、食品グループの推定結果の事後確率を基に決定しているが、詳細は割愛する。

食品グループを推定する対話管理はインタビューの対話管理とは別に行っている。対話管理は初めにシステムが推定した食品グループを質問し、ユーザに答えてもらう。もし、その中に正解がないとユーザが答えた場合正解の食品グループをユーザに聞く。この食品グループを推定する対話が終了するとインタビューに戻る。

4.5 潤滑発話候補の生成

潤滑発話も質問と同様にユーザ発話に対して答えられると判断された潤滑発話候補を対話知識の中からすべて選択し生成する。潤滑発話の生成はユーザの発話タイプとその発話内容に含まれる食品を判断して行っている。ユーザ発話が肯定と判断された場合否定的な応答以外の潤滑発話が生成される。例えば、自己開示:「私の好物です!」、共感:「いいですねー!」、ユーザへの質問:「美味しかったですか?」といった発話が候補に追加される。否定の場合は否定的な応答に対する潤滑発話

*1 CRFsuite, <http://www.chokkan.org/software/crfsuite/>.

表 2: 潤滑発話の発話数

発話タイプ	発話数
共感	26
量が多いことについて	22
量が少ないことについて	50
ユーザへの質問	6
自己開示	2
相槌	6
否定的な応答に対して	7
個々の食品に大して	323

話のみが生成される。例えば、「テンションが下がりますね・・・」、「代わりに好きなものを食べに行きましょうよ!」といった発話が候補に追加される。また、ユーザ発話内に特定の食品が含まれている場合には食品ごとに対応した潤滑発話が生成される。例えば、ユーザ発話に「納豆」と食品名が合った場合、「納豆は苦手です・・・」といった発話が候補に追加される。さらに、食事量についての内容があれば食事量に関する潤滑発話を生成する。例えば、ユーザが「食べすぎちゃいました」などと食事量が多い発話をした場合「健康に気をつけてくださいね!」といった発話が候補に追加される。このような様々な潤滑発話を収集するために、複数のユーザにアンケートを行った。このアンケートの結果収集した潤滑発話は 442 発話であり、内訳は表 2 のとおりである。

4.6 応答の決定

ユーザ発話に対して応答可能な質問と潤滑発話の候補が多数生成されるため、その中からシステムが行う応答を 1 つ選択する必要がある。まず、質問または潤滑発話どちらを生成するか判断を行う。対話システムは 1 つ前の質問をした後に応答した潤滑発話数 n を記録しており、その n が事前に規定した潤滑発話数 N 以上であれば、質問の応答候補を選択しそれ以下であれば潤滑発話の応答候補を選択する。

潤滑発話では、似たような発話を避けるために n 回の潤滑発話で同じ種類の発話候補を選択しないようにすでに発話した候補を排除する。例えば、1 回目の潤滑発話で自己開示の潤滑発話が選ばれた場合、2 回目以降の潤滑発話は自己開示以外の潤滑発話が候補となる。この候補の中からランダムに潤滑発話は選択される。

4.7 システムとユーザとの対話例

本稿で構築した対話システムとユーザとの対話例を記述する。例示する対話は 1 人のユーザが潤滑発話 3 回、1 回、なしの対話システムと対話したものである。

- 潤滑発話 3 回

S:昨日の昼食はどんなものを食べましたか? U:ハンバーグを食べました S:ハンバーグだったんですね U:はい、ハンバーグでした。 S:ハンバーグ好きなのですか? U:はい、大好きです。 S:実は私もハンバーグ食べましたよ。 U:ええっそうなんですか、おいしかったですか? S:昨日の昼食でハンバーグ以外に食べたものはありますか?

- 潤滑発話 1 回

S:昨日の昼食は何にしましたか? U:昼食はお好み焼きにしました。 S:マヨネーズはかけましたか? U:はい、たっぷりかけました。 S:昨日の昼食でスープとか汁物以外に食べたものはありますか?
--

表 3: 評価実験のアンケート

質問項目	
Q ₁	システム応答に意味がある(とても) ↔ システム応答に意味がない(とても)
Q ₂	非常に楽しい ↔ 非常に楽しくない
Q ₃	非常に自然な ↔ 非常に不自然な
Q ₄	非常に温かい ↔ 非常に冷たい
Q ₅	非常にまた話したい ↔ 非常に話したくない
Q ₆	非常に対話が弾む ↔ 非常に対話が弾まない
Q ₇	非常に簡潔な ↔ 非常に回りにくい

- 潤滑発話なし

S:昨日の昼食は何を食べました? U:お好み焼きを食べました。 S:昨日の昼食で他に食べたものはありますか?
--

5. 実験

潤滑発話の有効性を示すためにクラウドソーシングを用いて 100 人のワーカに 3 種類の対話システムを使ってもらって評価した。なお、実験データの解析はシステムの問題で途中で対話が終わってしまったワーカや全く同じ発話を繰り返しているワーカなどを除いた 92 人のワーカの対話に対して行った。

5.1 ベースライン手法

本実験では、潤滑発話ありとなしにおいてユーザの評価に差があるのか比較するため、従来の対話システムの手法である潤滑発話なしの対話システムをベースライン手法とした。また、潤滑発話数での比較として潤滑発話 3 回と 1 回の対話システムを用いた。潤滑発話 3 回は事前に行ったアンケート結果から最もユーザに支持されており、ユーザはなるべく長くシステムとの対話を望んでいるという知見が得られたため使用した。しかし、予備実験において、潤滑発話 3 回と潤滑発話なしを比較すると、潤滑発話なしとありに有意差が見られなかった。原因として、潤滑発話が増えたことでユーザ発話の理解度が下がったことと、潤滑発話数が多いと適切な対話制御ができないことがあげられる。この 2 つの問題を解決するために、潤滑発話 1 回の対話システムを用いた。潤滑発話の回数を減らすことで、ユーザ発話を理解する回数が減り精度の向上が見込まれる。また、潤滑発話を何度も繰り返す必要がないため潤滑発話が連続した場合の対話制御を考える必要がない。

5.2 実験方法の詳細

ユーザはクラウドソーシングの作業画面に記載してある URL から対話システムが起動されているサーバにアクセスする。アクセス後、挨拶を行い対話システムを起動する。対話システムは潤滑発話なし、1 回、3 回の中からランダムに選択されインタビューが始まる。1 つの対話システムとの対話が終わると作業画面で表 3 のような尺度の 5 段階評価のアンケートに答えてもらった。これをすべての対話システムとの対話が終わるまで繰り返した。

また、1 つの対話システムの最大対話ターン数は 33 ターンに設定しており、それ以上ユーザが発話を行うとその対話はそこで強制的に終了する。食事内容を聞く時間帯については朝、昼、夜の 3 つの時間を設定した。

5.3 実験結果

図 3 がそれぞれの尺度の平均値を取った結果であり、評価値が高いほど良い評価となっている。また、エラーバーがそれぞれの標準偏差を示している。全体的に潤滑発話ありのシステムの方が潤滑発話なしのシステムより良い評価を示している。また、潤滑発話なしと潤滑発話 1 回、3 回、潤滑発話 1 回と 3 回にウィルコクソンの順位和検定を行った。図 3 に * で示している。多くの尺度で潤滑発話ありとなしに優位差が見られた。

評価尺度を尺度ごとに評価値のヒストグラムとして表すと“自然さ”と“対話”が弾むの尺度において評価尺度“2”、“4”

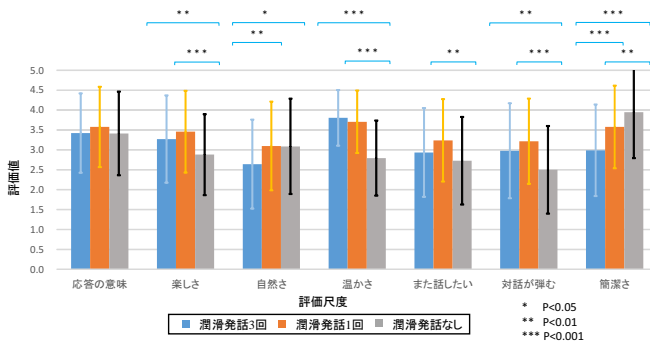


図 3: 評価の平均値

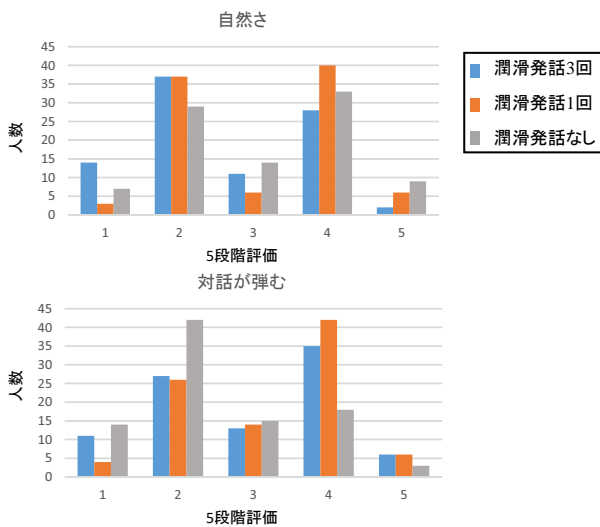


図 4: “自然さ”と“対話が弾む”尺度の評価値のヒストグラム

に特徴的な 2 つのグループができていた (図 4)。このことについては 5.4 節で原因を議論する。

100 人のユーザとシステムとの全ての対話からユーザの発話 1,000 個をランダムに抜き出し、理解できているかを調査した。ここでの正解基準は発話タイプ推定は人が確認した結果と一致したかどうかであり、意味情報抽出は抽出すべきすべてのユーザ情報が抽出された場合である。発話タイプ推定の正答率は 91.7% であり、意味情報抽出の正答率は 84.8% である。また、食品を推定した対話からランダムにシステム応答 200 個を抜き出し、システムが提示した候補内に正解の食品グループが含まれているか判定した。このときユーザが回答した食品グループを正解として判断するのではなく、アノテータが思った食品グループを正解とし判断した。例えば、「リンゴ」に対してユーザが食品グループを主食と答えていたとしても、アノテータがデザートと考えていれば、デザートがシステムの提示した候補内に入っていた場合を正解とした。結果として、84.0% の正答率で正解の食品グループを候補にあげることができていた。

5.4 議論

図 3 から、潤滑発話 1 回の方が対話の楽しさや温かさなどユーザに好印象を与える結果となっており、潤滑発話の有効性が示されている。しかし、潤滑発話 3 回では“自然さ”の尺度が潤滑発話なしの方が高くなっており、5% の有意差ができてしまっている。これは、予備実験の結果と同様に、潤滑発話はユーザがどのような内容を発話するのか予測がつかないため、うまく言語理解ができないことと潤滑発話はランダムに発話候

補の中から選択するため、1 つ前の潤滑発話の内容とは関係のない潤滑発話が生成されてしまうことが原因であると考えられる。そのため、潤滑発話の対話制御を考える必要のない潤滑発話 1 回のシステムでは“自然さ”の尺度は潤滑発話なしの対話システムとほぼ同じであり、自然な対話を実現することができている。

また、“自然さ”と“対話が弾む”の尺度において 2 つのグループが出来たことについてはユーザの感想から対話システムに求める期待値の差が原因ではないかと考えられる。自然さの尺度で“2”を選んだグループは、対話システムに対して「面白くない」や「あまりちゃんと答えられない」という感想が多い傾向が見られた。しかし、“4”を選んだグループは、「どのようにユーザ発話を与えれば理解してくれるのか考えるのが楽しい」や、「ロボットと話せるチャットが楽しい」など対話システムに対して人並みに理解してほしいという期待を求めている傾向が見られた。

6. おわりに

本稿では、潤滑発話を質問の合間に用いることとそれに伴う対話管理手法を提案した。この手法を使用した対話システムをユーザに使用してもらい評価実験を行った。実験の結果、潤滑発話の有効性が示された。

現在のシステムは潤滑発話の回数が固定されているが、今後はユーザの反応に合わせて、柔軟な回数と適切なタイミングで潤滑発話を生成できる対話アルゴリズムを検討していく。

参考文献

- [Fan 08] Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J.: LIBLINEAR: A Library for Large Linear Classification, *Journal of Machine Learning Research*, Vol. 9, (2008)
- [Gary 85] Gary, K., Steve, N., and John, M.: MORE: an intelligent knowledge acquisition tool, in *Proc. IJCAI*, Vol. 85 (1985)
- [Higashinaka 15] Higashinaka, R., Funakoshi, K., Araki, M., Tsukahara, H., Kobayashi, Y., and Mizukami, M.: Towards Taxonomy of Errors in Chat-oriented Dialogue Systems, in *Proc. SIGDIAL* (2015)
- [Johnston 13] Johnston, M., Ehlen, P., Conrad, F. G., Schober, M. F., Antoun, C., Fail, S., Hupp, A., Vickers, L., Yan, H., and Zhang, C.: Spoken dialog systems for automated survey interviewing, in *Proc. SIGDIAL* (2013)
- [Otsuka 13] Otsuka, T., Komatani, K., Sato, S., and Nakano, M.: Generating More Specific Questions for Acquiring Attributes of Unknown Concepts from Users, in *Proc. SIGDIAL* (2013)
- [Stent 06] Stent, A., Stenchikova, S., and Marge, M.: Dialog Systems for Surveys: the Rate-a-Course System., in *Proc. SLT* (2006)
- [東中 14] 東中: 雑談対話システムに向けた取り組み, 人工知能学会 言語・音声理解と対話処理研究会, Vol. 70, (2014)
- [磯村 09] 磯村, 鳥海, 石井: HMM による非タスク指向型対話システムの評価, 電子情報通信学会論文誌 D-92(4) (2009)